

Research Notes

Office of Research and Analysis

RN-25, March 2006

Test Development and Technical Information on the Writing Section of the SAT Reasoning Test $^{\mathsf{TM}}$

Jennifer L. Kobrin and Ernest W. Kimmel

Introduction and Description of the SAT® Writing Section

The new SAT Reasoning Test[™], first administered in March 2005, was designed to incorporate a number of important changes. Perhaps the most noteworthy change to the test was the addition of a writing section. Writing is a core skill needed for success in both college and the workplace. Table 1 summarizes the content of the SAT[®] writing section, which includes multiple-choice questions and an essay.

Multiple-Choice Questions

The multiple-choice questions assess how well students use standard written English and test students' ability to identify

Table 1

Content of the SAT Writing Section					
Content	Number of Questions	Time	Scores		
Essay	1	25 minutes	0, 2–12 (0 if illegible or off topic)		
Improving Sentences (Multiple Choice)	25	One 25- minute			
Identifying Sentence Errors (Multiple Choice)	18	multiple- choice section, and	20-80		
Improving Paragraphs (Multiple Choice)	6	one 10-minute multiple- choice section			
Total	50 questions	60 minutes	200-800		

sentence errors, improve sentences, and improve paragraphs. The multiple-choice writing questions are used to assess students' ability to:

- Use language that is consistent in tenses and pronouns
- Understand parallelism, noun agreement, and subjectverb agreement
- Understand how to express ideas logically
- Avoid ambiguous and vague pronouns, wordiness, improper modification, and sentence fragments
- Understand proper coordination and subordination, logical comparison, diction, idiom, modification, and word order

The multiple-choice writing questions do not ask students to define or use grammatical terms and do not test spelling or capitalization. The improving sentences questions ask students to recognize and correct faults in grammar and sentence structure and to recognize effective sentences that follow the conventions of standard written English. The identifying sentence errors questions ask students to recognize and correct errors in grammar and usage and to recognize effective sentences that follow the conventions of standard written English. The improving paragraphs questions ask students to apply the conventions of standard written English while editing and revising sentences in the context of a paragraph or an entire essay, or while organizing and developing paragraphs in a coherent and logical manner.

Essay Question

The SAT writing section provides 25 minutes for students to write a first-draft essay in response to an assignment question. Students are presented with a short paragraph adapted from a published text that offers a perspective on an issue, and an assignment question that asks for the students' point

of view. Students must think critically about the issue and develop their point of view, using reasoning and examples taken from their reading, studies, experiences, or observations to support their ideas. The essay measures students' ability, under timed conditions, to do the kind of writing required in most college courses—writing that emphasizes precise use of language, logical presentation of ideas, development of a point of view, and clarity of expression. SAT essay prompts are developed according to the following guidelines:

- They should be accessible to the general test-taking population, including students for whom English is not a first or best language.
- They should be relevant to a wide range of fields and interests, and neither require specialized knowledge nor give
 an advantage to students who have completed a specific
 course of study.
- They should engage high-school-age students while stimulating critical reflection about important topics.
- They should be free of figurative or technical language or specific literary references.
- They should give students the opportunity to use a broad spectrum of experiences, learning, and ideas to support their point of view.

A study was conducted to investigate the impact of the type of essay prompt used in the SAT on ethnic, language, and gender groups (Breland, Kubota, Nickerson, Trapani, and Walker, 2004). The results of this study indicated that the type of essay prompt used in the SAT did not disadvantage any particular group of students. Further studies will be conducted as more essay prompts are tested.

Development of the SAT Writing Section

An important goal in revising the SAT was to strengthen the links between the skills measured by the SAT Reasoning Test, high school and college curricula, and instructional practice. The design of the writing section was informed by long experience with the English Composition Test and its successor, the SAT Subject Test in Writing. In addition, several groups of high school and college educators helped to develop and define the new SAT writing section. The College Board's Test Development Committee in Writing continues to monitor item development and student performance on the writing section.

Early in the development of the SAT writing section, a survey of English and language arts teachers was conducted to determine the frequency with which specific reading and writing skills were covered in the classroom and how important teachers felt that these skills were for students entering their first year of college (Milewski, Johnsen, Glazer, and Kubota, 2005). Data were collected via a survey Web site and through paper survey instruments. Of the 2,351 teachers who responded, 1,044 taught at the college level and 1,307 taught at the high school level. Sixty-five percent of the respondents were women.

Overall, this survey study confirmed that the design of the SAT writing section was aligned with those writing skills that teachers deemed to be important and that were included in the taught curriculum. Table 2 presents the alignment between the items in the writing section and the writing process skills identified in the survey.

Not all of the skills deemed important by teachers can be assessed in a standardized testing context. The skills not measured by the SAT writing section are:

Table 2Alignment Between Writing Process Skills and SAT Writing Questions

Writing Process Skill	Essay Prompt	Improving Paragraphs
Writing personal narratives	X	
Using literal and figurative language appropriately	X	X
Using sentence variety	X	X
Demonstrating insight and/or creativity in the writing task	X	
Using topic sentences	X	X
Using appropriate voice, tone, and style	X	X
Focusing on a purpose for writing	X	
Writing persuasive and/or argumentative essays	X	
Organizing paragraphs and using appropriate transitions	X	X
Writing effective introductions and conclusions	X	X
Using writing and reading as tools for critical thinking	X	
Developing a logical argument	X	
Writing a unified essay	X	X
Using supporting details and examples	X	
Writing a clear and coherent essay	X	X

Source: Milewski, et al. (2005, Table 8).



- · Creative writing
- Using peer groups for feedback and revision
- Responding to the needs of different audiences
- · Using prewriting techniques to generate texts
- Generating multiple drafts while creating and completing texts
- Understanding writing as a process of invention and rethinking
- · Learning strategies for revising, editing, and proofreading
- Understanding the purposes of different forms of writing
- · Writing analyses and evaluations of texts

The inclusion of an essay as part of the SAT writing section is an important component in assessing writing process skills. The seven skills receiving the most attention in the classroom are all assessed directly by the essay portion of the test. One of the multiple-choice question types—improving paragraphs—also taps into students' knowledge of writing process skills. Many of the more discrete skills related to grammar, usage, and sentence structure are assessed by the multiple-choice questions of the SAT writing section. The alignment of the test questions with these skills that contribute to effective writing are shown in Table 3.

Only two common skills are not tested by the multiple-choice questions: controlling errors in spelling and using all punctuation appropriately. However, because the holistic scoring approach to the essay requires that all features of a written response be taken into account, the evaluation of students' essays may be affected by their command of these two skills.

Scoring of the SAT Writing Section

The multiple-choice raw score accounts for about 75 percent of the total scaled score, and the essay for about 25 percent. The raw score on the essay is the sum of the two scores received from the different raters. The multiple-choice raw score is arrived at in the same way as the other multiple-choice sections on the SAT: +1 point per right answer, -1/4 point per wrong answer. Omitted questions do not affect the score. The scaled score ranges from 200 (lowest) to 800 (highest).

The elements of writing that can be assessed through the essay are reflected in the scoring guide that is used by readers to evaluate and score the student essays holistically. The scoring guide is structured on the six-point scale pre-

Table 3

Alignment Between Grammar, Usage, and Sentence Structure Skills and the Problems Tested by SAT Writing Questions

Grammar, Usage, and Sentence Structure Skills	Improving Sentences	Identifying Sentence Errors	Improving Paragraphs
Avoiding faulty predication in sentences	X	X	X
Avoiding dangling modifiers	X		
Using comparative modifiers appropriately	X	X	
Using appropriate idiomatic words, phrases, or structures	X	X	X
Avoiding weak, passive constructions	X		
Using connectives appropriately	X	X	X
Avoiding illogical comparisons	X	X	
Subordinating and coordinating ideas in sentences	X	X	X
Avoiding pronoun shift	X	X	X
Combining sentences appropriately			X
Maintaining parallel structure in sentences	X	X	X
Using appropriate verb forms	X	X	X
Avoiding wordiness	X	X	X
Controlling errors in subject-verb agreement	X	X	
Avoiding errors in pronoun agreement, case, and reference	X	X	
Maintaining tense sequences	X	X	X
Making acceptable word choices	X	X	X
Avoiding run-on sentences	X		
Avoiding sentence fragments	X		X
Avoiding comma splices	X		X

sented in Table 4. In holistic scoring, a piece of writing is considered as a total work, the whole of which is greater than the sum of its parts. The essay is scored by qualified readers who take into account such aspects as complexity of thought, substantiality of development, and facility with language. A reader will not judge a work based on its separate traits, but rather on the total impression it creates.

Each essay is scored independently by two qualified readers, on a scale of 1 to 6 by each reader, with the combined score for both readers ranging from 2 to 12. (Essays not written on the essay assignment will receive a score of zero.) If the two readers' scores differ by more than one point, a third reader will score the essay and his or her score is doubled.

Essays for the new SAT writing section are scanned and distributed to readers via the Web. Readers are supervised

Table 4

Scoring Guide for SAT Essay		
SCORE OF 6	SCORE OF 5	SCORE OF 4
An essay in this category demonstrates clear and consistent mastery, although it may have a few minor errors. A typical essay: • effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking, using clearly appropriate examples, reasons, and other evidence to support its position • is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas • exhibits skillful use of language, using a varied, accurate, and apt vocabulary • demonstrates meaningful variety in sentence structure • is free of most errors in grammar, usage, and mechanics	An essay in this category demonstrates reasonably consistent mastery, although it will have occasional errors or lapses in quality. A typical essay: • effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons, and other evidence to support its position • is well organized and focused, demonstrating coherence and progression of ideas • exhibits facility in the use of language, using appropriate vocabulary • demonstrates variety in sentence structure • is generally free of most errors in grammar, usage, and mechanics	An essay in this category demonstrates adequate mastery, although it will have lapses in quality. A typical essay: • develops a point of view on the issue and demonstrates competent critical thinking, using adequate examples, reasons, and other evidence to support its position • is generally organized and focused, demonstrating some coherence and progression of ideas • exhibits adequate but inconsistent facility in the use of language, using generally appropriate vocabulary • demonstrates some variety in sentence structure • has some errors in grammar, usage, and mechanics
SCORE OF 3	SCORE OF 2	SCORE OF 1
An essay in this category demonstrates developing mastery, and is marked by ONE OR MORE of the following weaknesses: • develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons, or other evidence to support its position • is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas • displays developing facility in the use of language, but sometimes uses weak vocabulary or inappropriate word choice • lacks variety or demonstrates problems in sentence structure • contains an accumulation of errors in grammar, usage, and mechanics	An essay in this category demonstrates little mastery, and is flawed by ONE OR MORE of the following weaknesses: • develops a point of view on the issue that is vague or seriously limited, and demonstrates weak critical thinking, providing inappropriate or insufficient examples, reasons, or other evidence to support its position • is poorly organized and/or focused, or demonstrates serious problems with coherence or progression of ideas • displays very little facility in the use of language, using very limited vocabulary or incorrect word choice • demonstrates frequent problems in sentence structure • contains errors in grammar, usage, and mechanics so serious that meaning is somewhat obscured	An essay in this category demonstrates very little or no mastery, and is severely flawed by ONE OR MORE of the following weaknesses: • develops no viable point of view on the issue, or provides little or no evidence to support its position • is disorganized or unfocused, resulting in a disjointed or incoherent essay • displays fundamental errors in vocabulary • demonstrates severe flaws in sentence structure • contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning

Essays not written on the essay assignment will receive a score of zero.

online by scoring leaders, who are experienced essay readers with special training in online scoring. By working with the readers via the Web, the College Board is able to attract a larger reader pool from across the country. The readers who score the SAT essay are experienced high school teachers and college faculty members who teach courses that require a substantial amount of writing. To qualify, readers must:

- Hold a bachelor's degree or higher
- Teach or have taught within the last five years a highschool- or college-level course that requires writing
- Have taught for at least a three-year period
- Reside in the continental United States, Alaska, or Hawaii

 Be a U.S. citizen, resident alien, or be authorized to work in the United States

For the March–June 2005 administrations, 85 percent of the essay scorers possessed an advanced degree, and 71 percent had taught for five years or more.

The readers are required to complete a rigorous online training course that familiarizes them with the principles of holistic scoring and teaches them to evaluate essays according to the agreed-upon standards. It should be noted that SAT readers are required to possess more experience than the readers of any state writing assessment. The training process is ongoing and does not end once a reader has qualified. The accuracy and fairness of the readers are evaluated regularly and frequently



by having readers successfully score a small set of prescored "calibration" essays. Calibration papers are used to clarify issues and provide feedback to readers once the scoring has begun.

To maintain reader accuracy throughout the scoring process, "validity papers"—clear examples of score points—are interspersed with other student responses. Scoring leaders review readers' scoring of selected essays, including validity essays, and provide feedback via phone and the Web when appropriate. Web-based scoring enables leaders to monitor readers in real time, informed by extensive real-time and summary reports on inter-rater reliability, validity, and calibration statistics. This monitoring and training program ensures the highest quality of scoring performance.

Preliminary Technical Information on the SAT Writing Section

This section presents statistical and psychometric information on the SAT writing section based on the first six administrations of the new test. Technical information for the SAT is typically reported to the public on an annual basis based on scores from the previous year's cohort of graduating seniors (referred to as college-bound seniors). For example, the 2005 College-Bound Seniors report shows the most recent SAT scores for students graduating high school in 2005. The first aggregate data on the writing section for an entire cohort will be reported in the 2006 College-Bound Seniors report. Percentiles are also based on the previous year's cohort. For example, the percentile reported for a math score on the March 2005 administration is based on the percentile for that score among all college-bound seniors in 2004. Because an entire cohort did not take the new writing section, percentiles will not be reported until data for one

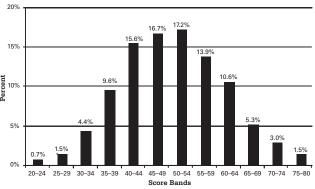


Figure 1. Multiple-choice score distribution for March–December 2005 administrations.

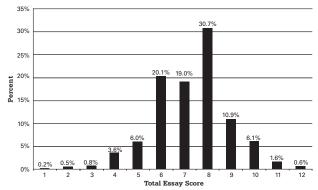


Figure 2. Essay score distribution for March–December 2005 administrations.

full cohort are available. However, an interim percentile table is currently available to provide some guidance in interpreting the composite writing score (see www.collegeboard.com/prod_downloads/student/testing/sat/scores/SATpercentile.pdf). Please note that the information presented in this section may change when data for the full cohort are available.

A total of 1,258,016 students took the SAT in the March–December 2005 administrations. The average score on the writing section was 502 with a standard deviation of 108. Figure 1 shows the distribution of multiple-choice writing scores for the March–December 2005 administrations. The average multiple-choice score was 50.2 with a standard deviation of 10.9. Figure 2 shows the distribution of essay scores for the March–December 2005 administrations. The distribution was essentially normal, with a mean score of 7.3, a median score of 7.0, and a modal score of 8.0.

Figure 3 shows the essay reader agreement statistics for March–December 2005. For the majority of the essays (56 percent), both readers assigned the same scores. There was a one-point difference in scores for 38 percent of the essays, and a two-point difference in scores for 6 percent of the essays. Less than one-half of a percent of the essays were assigned scores that differed by three points or more.

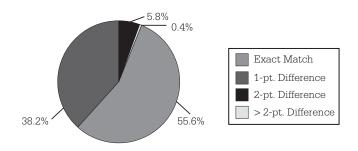


Figure 3. Essay scoring agreement by first two readers for March–December 2005 administrations.

Reliability of the Writing Section

For the March–December 2005 administrations, scores on the writing section correlated .84 with critical reading scores and .72 with mathematics scores, indicating a high degree of correspondence among the three sections of the SAT.

Multiple-Choice and Composite Reliability

For the first seven forms administered in March, May, and June 2005, internal consistency reliability estimates for the writing multiple-choice questions ranged from .88 to .90, with a mean of .89 across forms. These estimates are based on the 49 multiple-choice items in the writing section. These reliability estimates are very similar to those for the now-defunct SAT Subject Test in Writing (.86–.92). The reliability estimates for the composite writing score (multiple-choice and essay questions combined) will be available soon on the College Board Web site.

Essay Reliability

A special study was conducted to produce preliminary reliability estimates for the essay component of the SAT writing section (Allspach and Walker, 2005). In this study, the same students wrote on two different essay topics within two weeks (in order to control for effects due to time and learning). Approximately 3,500 non-ESL juniors in 45 schools participated in the study. Raters received similar training to those rating the operational SAT essay responses. Three reliability estimates were calculated:

- *Inter-rater reliability*—the correlation between scores from two raters scoring the same essay.
- Essay scoring reliability—the correlation between average scores from two sets of raters scoring the same essay. This represents the consistency in the scoring method itself.
- Essay observed score reliability—the correlation between average scores from two sets of raters scoring different essays. This represents the proportion of true score variance in the essay score itself, and is the relevant coefficient to the use of essay scores to estimate examinees' writing ability.

Table 5 presents the reliability estimates. The estimated true score reliability across all prompts is .76. The average

Table 5

SAT Essay Reliability Estimates						
Prompt	Inter-Rater Reliability	Essay Scoring Reliability	Essay Observed Score Reliability			
A	.81	.89	.68			
В	.77	.87	.67			
С	.78	.88	.66			
D	.80	.89	.67			

alternate-forms reliability coefficient across the four prompts that was used to compute the standard error of measurement for the writing section was .6679.

Based on these statistics, the standard error of measurement for the writing section is 1.04, as shown by the formula below:

$$SEM = sd\sqrt{1 - r_{xx}}$$

This means that for an essay subscore of 8, there is a 68 percent probability that the student's true score is between 6.96 and 9.04.

Predictive Validity of the Writing Section

The College Board commissioned the American Institutes for Research to conduct a study to investigate the validity of the SAT writing section for predicting both first-year college grade point average (FGPA) and for predicting course grades in college English composition courses (Norris, Oppler, Kuang, Day, and Adams, in press). The study focused on the extent to which the new SAT writing section predicted college performance after SAT verbal and mathematics scores and high school grade point average (HSGPA) were already considered.

To conduct this study, an experimental version of the new SAT writing section was administered to samples of incoming first-year students at 13 colleges and universities around the United States during summer orientation or in the initial weeks of the fall 2003 semester. At the completion of the spring 2004 semester, information regarding the academic performance of student participants during their first year in college was obtained (i.e., FGPAs and grade point averages in English composition courses [ECGPA]). The resulting participant sample consisted of 1,572 students who took the prototype SAT writing section and who were also identified in the College Board database (i.e., who also took the SAT Reasoning Test).

¹ The reliability estimates for the October and December 2005 forms will be available in spring 2006.



Table 6

Weighted-Average Correlations for all Predictors with First-Year College GPA and English Composition GPA

		0	0		-	
	FGPA			ECGPA		
Predictor	N	Corrected	Uncorrected	N	Corrected	Uncorrected
SAT-V	1,248	.49	.32	891	.30	.20
SAT-M	1,248	.47	.29	891	.23	.10
SAT-T	1,248	.51	.35	891	.28	.17
SAT-ES	1,248	.20	.16	891	.18	.14
SAT-MC	1,248	.45	.30	891	.31	.22
SAT-W	1,248	.46	.32	891	.32	.24
HSGPA	1,248	.43	.38	891	.35	.32

Note: Corrected for multivariate range restriction (Lord and Novick, 1968). SAT-V = SAT verbal; SAT-M = SAT mathematics; SAT-T = SAT combined (verbal + mathematics); SAT-ES = SAT writing essay; SAT-MC = SAT writing multiple choice; SAT-W = SAT writing composite (essay + multiple choice); HSGPA = high school grade point average.

Correlations were computed between each predictor (SAT verbal, mathematics, and writing scores, and HSGPA) and each criterion variable (FGPA and ECGPA). In addition, hierarchical regression analyses were used to assess the incremental validity results for various combinations of predictors and FGPA. All analyses were conducted within each institution, and weighted averages were calculated across institutions to derive overall study estimates. Statistical procedures to correct for multivariate range restriction (Lord and Novick, 1968) and shrinkage (Rozeboom, 1978) were applied.

As shown in Table 6, the average corrected validity coefficient for FGPA ranged from .20 for the essay component of the SAT writing section score to .51 for the SAT combined score. In fact, when corrected for range restriction, all of the predictors except the SAT writing essay had validity coefficients of .43 or greater. With respect to ECGPA, Table 6 shows that the average corrected validity coefficient ranged from .18 for the SAT writing essay to .35 for HSGPA. Table 6 also shows that the combined SAT writing section, the SAT writing multiple-choice, and the SAT verbal scores were also fairly predictive of ECGPA with corrected validity coefficients of .32, .31, and .30, respectively.

As shown in Table 7, the incremental validity of the writing scores when added to verbal scores, mathematics scores, and HSGPA was .01 when corrections for range restriction and shrinkage were made. The multiple correlation for this fully corrected model was .60, providing the best estimate of the validity of SAT verbal, mathematics, and writing scores, along with HSGPA, for predicting FGPA.

Table 7

Weighted-Average Incremental Validity Results Across Institutions for Predicting FGPA

		Adjusted		Unadjusted	
	Step	R	ΔR	R	ΔR
Corrected	1: SAT-V + SAT-M + HSGPA	.59		.63	
	2: SAT-V + SAT-M + HSGPA + SAT-W	.60	.01	.64	.02
Uncorrected	1: SAT-V + SAT-M + HSGPA	.46		.51	
	2: SAT-V + SAT-M + HSGPA + SAT-W	.47	.01	.53	.02

Note: N=1,248. Corrected correlations were corrected for range restriction. Adjusted correlations were adjusted for shrinkage (Rozeboom, Formula 8).

The complete report, available soon at www.collegeboard.com/research/home, presents the results of regression models designed to assess the incremental validities of various combinations of the predictors of interest, as well as individual results for the 13 participating institutions.

Conclusions

Based on statistics from the first few administrations of the new SAT writing section, the test is performing as expected. The reliability of the writing section is very similar to that of other writing assessments. Based on preliminary validity research, the writing section is expected to add modestly to the prediction of college performance when critical reading and mathematics scores are considered. A full validity study on the new SAT is being planned for the cohort of students who will complete their first year of college in 2007.

Jennifer L. Kobrin is a research scientist at the College Board. Ernest W. Kimmel is an educational measurement consultant based in Hopewell, New Jersey.

References

Allspach, J. R., & Walker, M. E. (2005). *Essay Reliability Coefficients*. ETS Memorandum, September 30, 2005.

Breland, H., Kubota, M., Nickerson, K., Trapani, C., & Walker, M. (2004). *New SAT writing prompt study: Analyses of group impact and reliability*. (College Board Research Report No. 2004-1). New York: The College Board.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Test Development and Technical Information on the Writing Section of the SAT Reasoning Test

Milewski, G. B., Johnsen, D., Glazer, N., & Kubota, M. Y. (2005). A survey to evaluate the alignment of the new SAT writing and critical reading sections to curricula and instructional practices. (College Board Research Report No. 2005-1). New York: The College Board.

Norris, D., Oppler, S., Kuang, D., Day, R., & Adams, K. (in press). *The College Board SAT writing validation study: An assessment of predictive and incremental validity.* (College Board Research Report 2006-2). New York: The College Board.

Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlation: A clarification. *Psychological Bulletin*, 85(6), 1348–51.

Office of Research and Analysis The College Board 45 Columbus Avenue New York, NY 10023-6992 212 713-8000

The College Board: Connecting Students to College Success

The College Board is a not-for-profit membership association whose mission is to connect students to college success and opportunity. Founded in 1900, the association is composed of more than 5,000 schools, colleges, universities, and other educational organizations. Each year, the College Board serves seven million students and their parents, 23,000 high schools, and 3,500 colleges through major programs and services in college admissions, guidance, assessment, financial aid, enrollment, and teaching and learning. Among its best-known programs are the SAT®, the PSAT/NMSOT®, and the Advanced Placement Program® (AP®). The College Board is committed to the principles of excellence and equity, and that commitment is embodied in all of its programs, services, activities, and concerns. For further information, visit www.collegeboard.com.

© 2006 The College Board. All rights reserved. College Board, Advanced Placement Program, AP, SAT, and the acorn logo are registered trademarks of the College Board. connect to college success and SAT Reasoning Test are trademarks owned by the College Board. PSAT/NMSQT is a registered trademark of the College Board and National Merit Scholarship Corporation. All other products and services may be trademarks of their respective owners. Visit the College Board on the Web: www.collegeboard.com.

Permission is hereby granted to any nonprofit organization or institution to reproduce this report in limited quantities for its own use, but not for sale, provided that the copyright notice be retained in all reproduced copies as it appears in this publication.

